## 空气质量指标说明

AQI：空气质量指数，其分级计算参考的标准是新的环境空气质量标准（GB3095-2012），参与评价的污染物为SO2、NO2、PM10、PM2.5、O3、CO等六项

PM10：沙尘暴主要物质

PM2.5：雾霾主要物质

## 分工

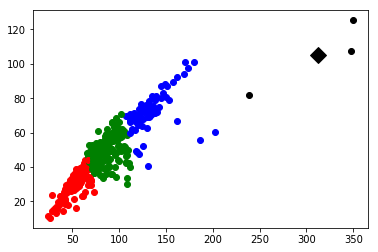
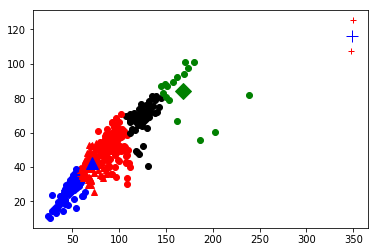
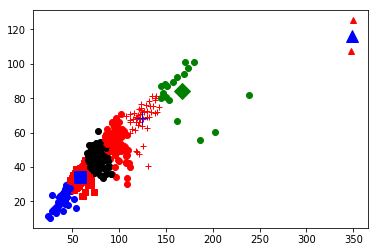
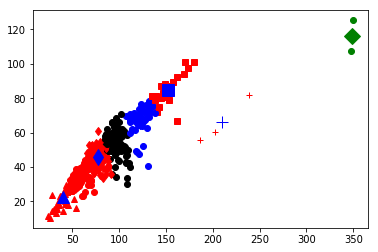
1. 李朝阳：组长，爬取全国空气质量数据，可视化
2. 米鹏：ppt制作。Weka聚类分析。爬取兰州天气数据。
3. 吴传曼、邬晋宁：论文撰写，数据可视化。

## 遇到的问题和期望的收获

### 问题

1. 由于数学功底不行，得到数据后并未进行有效的统计
2. R语言基础不行，可视化步骤做的不完善
3. 得出的结论很牵强。
4. 由于环境问题，并未在Hadoop上运行kmeans算法。
5. 课堂讲解时并未将清本小组思路。
6. 对全国300多城市各个空气指标进行了分年的平均。想要根据PM10，PM2.5两个维度聚类，得出有的城市主要污染物是PM10（比如兰州），有的城市主要污染物则是PM2.5（比如河北）。

然而不论如何改变K值，聚类结果却大都沿Y=x这条直线。（详见代码 /Kmeans和可视.ipynb）



### 收获

1. 理解了聚类算法的基本思想和原理。
2. 初步分析了空气质量数据。验证了空气质量确实越来越差。